

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-334115

(43)Date of publication of application : 18.12.1998

(51)Int.Cl.

G06F 17/30

(21)Application number : 09-146259

(71)Applicant : FUJI XEROX CO LTD

(22)Date of filing : 04.06.1997

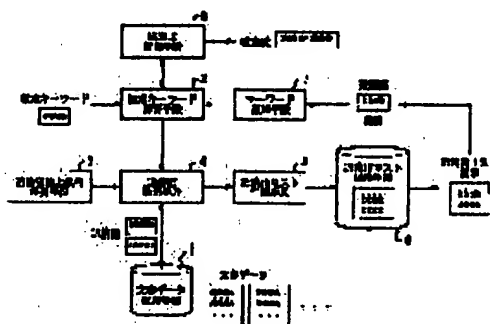
(72)Inventor : KURAMOCHI TSUTOMU  
KURAHASHI MASAYUKI

## (54) RETRIEVAL EXPRESSION GENERATING DEVICE

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide the retrieval expression generating device which lightens the burden of retrieval expression generation on a user by efficiently displaying information for narrowing down requests to retrieve documents.

**SOLUTION:** When this retrieval expression generating device generates a retrieval expression for document retrieval, a document data holding means 1 holds pieces of document data. Further, a retrieval key word holding means 2 holds an inputted retrieval key word. A near word extraction condition holding means 3 holds inputted near word extraction conditions. A near word extracting means 4 extracts near words from the pieces of document data according to the retrieval key word and near word extraction conditions. A near word list generating means 5 generates a new word list from the near words and a near word list holding means 6 holds it. A key word adding means 7 accepts choices to the near words included in the displayed near word list and adds the selected near words as retrieval key words. A retrieval expression generating means 8 generates a retrieval expression from the retrieval key words.



## LEGAL STATUS

[Date of request for examination]

10.07.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the  
examiner's decision of rejection or application converted  
registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of  
rejection][Date of requesting appeal against examiner's decision of  
rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-334115

(43) 公開日 平成10年(1998)12月18日

(51) Int.Cl.<sup>8</sup>

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/403

15/40

15/403

3 3 0 A

3 7 0 A

3 2 0 D

3 5 0 C

審査請求 未請求 請求項の数 3 O L (全 8 頁)

(21) 出願番号

特願平9-146259

(22) 出願日

平成9年(1997)6月4日

(71) 出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂二丁目17番22号

(72) 発明者 倉持 勉

神奈川県足柄上郡中井町境430 グリーン

テクなかい 富士ゼロックス株式会社内

(72) 発明者 倉橋 政之

神奈川県足柄上郡中井町境430 グリーン

テクなかい 富士ゼロックス株式会社内

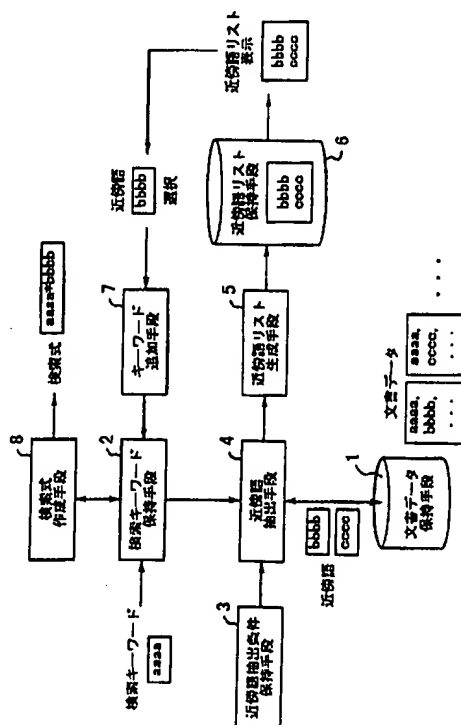
(74) 代理人 弁理士 服部 毅巖

(54) 【発明の名称】 検索式作成装置

(57) 【要約】

【課題】 文書の検索要求を絞り込むための情報提示を効率良く行い、検索式作成にかかるユーザの負担を軽減した検索式作成装置を提供すること。

【解決手段】 本発明の検索式作成装置にて文書検索のための検索式を作成する場合、文書データ保持手段1は複数の文書データを保持する。また、検索キーワード保持手段2は入力される検索キーワードを保持する。近傍語抽出条件保持手段3は入力される近傍語抽出条件を保持する。近傍語抽出手段4は検索キーワード及び近傍語抽出条件に基づいて、複数の文書データから近傍語を抽出する。近傍語リスト生成手段5は近傍語から近傍語リストを生成し、近傍語リスト保持手段6がこれを保持する。キーワード追加手段7は表示された前記近傍語リストに含まれた近傍語への選択を受け付け、選択された近傍語を検索キーワードとして追加する。検索式作成手段8は検索キーワードから検索式を作成する。



## 【特許請求の範囲】

【請求項1】 文書検索のための検索式を作成する検索式作成装置において、

複数の文書データを保持する文書データ保持手段と、  
入力される検索キーワードを保持する検索キーワード保持手段と、

入力される近傍語抽出条件を保持する近傍語抽出条件保持手段と、

前記検索キーワード及び前記近傍語抽出条件に基づいて、前記複数の文書データから近傍語を抽出する近傍語抽出手段と、

前記近傍語から近傍語リストを生成する近傍語リスト生成手段と、

前記近傍語リストを保持する近傍語リスト保持手段と、  
表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を検索キーワードとして追加するキーワード追加手段と、

前記検索キーワードから検索式を作成する検索式作成手段と、

を有することを特徴とする検索式作成装置。

【請求項2】 前記近傍語リスト生成手段は、前記近傍語を品詞別に分類して前記近傍語リストを生成することを特徴とする請求項1記載の検索式作成装置。

【請求項3】 複数の文書データから文書を検索する文書検索装置において、

複数の文書データを保持する文書データ保持手段と、  
入力される検索キーワードを保持する検索キーワード保持手段と、

入力される近傍語抽出条件を保持する近傍語抽出条件保持手段と、

前記検索キーワード及び前記近傍語抽出条件に基づいて、前記複数の文書データから近傍語を抽出する近傍語抽出手段と、

前記近傍語から近傍語リストを生成する近傍語リスト生成手段と、

前記近傍語リストを保持する近傍語リスト保持手段と、  
表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を検索キーワードとして追加するキーワード追加手段と、

前記検索キーワードから検索式を作成する検索式作成手段と、

前記検索式にて前記文書データから文書を検索する検索手段と、

を有することを特徴とする文書検索装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は検索式作成装置及び文書検索装置に関し、特に文書検索のための検索式を作成する検索式作成装置及び複数の文書データから文書を検索する文書検索装置に関する。

## 【0002】

【従来の技術】 多量の文書データから必要に応じて文書データを検索する方法には、シソーラスに基づいて統制されたキーワードや分類コード等の情報を利用する方法のほかに、文書全文を直接参照して検索する全文検索（フルテキストサーチ）方法がある。前者の方法に比べて後者の方法では検索もれが少ない。

【0003】 しかし、全文検索方法では検索もれが少ない代わりに多量の文書がヒットとなってしまい、この内容を逐一チェックするには大変な労力を要する。例えば、「有害物質の輸出入の規制」に関する内容の文書を検索する場合に、「有害物質」、「輸出入」、「規制」という単語を含むことを検索条件としてしまうと、これらの単語が散在した文書もヒットとなってしまふ場合がある。

【0004】 そこで、この全文検索方法で情報をより正確に絞り込むために、ユーザが指定した検索語（検索キーワード）の近傍のテキスト情報を利用する方法が提案されている。

【0005】 例えば、検索者が2つの検索キーワード及びその位置関係を指定すると、その検索条件に合致する検索結果を出力する「文書検索方法および装置」が、特開平4-293161号公報に示されている。

【0006】 この方法では、上記で例とした「有害物質の輸出入の規制」に関する内容の文書を抽出する場合、「有害物質」という単語と「輸出入」という単語とを含み、その間隔は2語以内であり、かつ、「輸出入」という単語と「規制」という単語とを含み、その間隔は2語以内である」ことを検索条件とすることができる。よって、検索キーワードが散在した文書のヒットを防止することができる。

【0007】 この方法は、検索要求が「有害物質の輸出入の規制」というように明確である場合には有効である。しかし、例えばユーザの検索要求で明らかになっているのが、検索キーワード「規制」のみ、という場合もある。

【0008】 このように、検索要求が定かでない場合に、ユーザが1つの検索キーワードを指定すると、その検索キーワードが出現する箇所の前後のテキストを抜き出してユーザに提示する「対話式データベース探索方法および装置」が特開平5-189487号公報に示されている。

【0009】 この方法では、例えば「規制」という検索キーワードを指定すると「有害物質の輸出入の規制」を始め、「～の規制」、「～に関する規制」といったテキストが列挙され、ユーザはその中から必要な情報を選択的に得ることができる。

## 【0010】

【発明が解決しようとする課題】 しかし、上記の方法では提示されるテキストが大量となり、ユーザの選択作業

にかかる負担が増大してしまうという問題点があった。なお、この原因の1つは、意味が殆ど同じであっても言い回しの異なるテキストは、すべて区別して提示してしまうことにある。例えば、「排気ガスの規制」と「排気ガスに関する規制」などがこれにあたる。

【0011】本発明は以上のような点に鑑みてなされたものであり、文書の検索要求を絞り込むための情報提示を効率良く行い、検索式作成にかかるユーザの負担を軽減した検索式作成装置を提供することを目的とする。

#### 【0012】

【課題を解決するための手段】本発明では上記課題を解決するために、文書検索のための検索式を作成する検索式作成装置において、複数の文書データを保持する文書データ保持手段と、入力される検索キーワードを保持する検索キーワード保持手段と、入力される近傍語抽出条件を保持する近傍語抽出条件保持手段と、前記検索キーワード及び前記近傍語抽出条件に基づいて、前記複数の文書データから近傍語を抽出する近傍語抽出手段と、前記近傍語から近傍語リストを生成する近傍語リスト生成手段と、前記近傍語リストを保持する近傍語リスト保持手段と、表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を検索キーワードとして追加するキーワード追加手段と、前記検索キーワードから検索式を作成する検索式作成手段と、を有することを特徴とする検索式作成装置が提供される。

【0013】このような検索式作成装置にて文書検索のための検索式を作成する場合、文書データ保持手段が複数の文書データを保持する。また、検索キーワード保持手段が入力される検索キーワードを保持する。近傍語抽出条件保持手段は入力される近傍語抽出条件を保持する。近傍語抽出手段は検索キーワード及び近傍語抽出条件に基づいて、複数の文書データから近傍語を抽出する。近傍語リスト生成手段は近傍語から近傍語リストを生成する。近傍語リスト保持手段は、生成された近傍語リストを保持する。キーワード追加手段は表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を検索キーワードとして追加する。そして検索式作成手段は検索キーワードから検索式を作成する。

【0014】このようにして文書検索のための検索式を作成すると、検索要求を絞り込むための情報、すなわち検索キーワード及び近傍語抽出条件に基づいて抽出される近傍語の提示が効率良く行われ、この近傍語を検索式を作成するための検索キーワードとして追加できるので、検索式作成にかかるユーザの負担を軽減することが可能となる。

#### 【0015】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照して説明する。図1は本発明の検索式作成装置の原理構成を示したブロック図である。

【0016】本発明の検索式作成装置は、文書データ保持手段1と、検索キーワード保持手段2と、近傍語抽出条件保持手段3と、近傍語抽出手段4と、近傍語リスト生成手段5と、近傍語リスト保持手段6と、キーワード追加手段7と、検索式作成手段8と、から構成される。

【0017】ここで、文書データ保持手段1は、複数の文書データを保持する。また、検索キーワード保持手段2は、入力される検索キーワードを保持する。近傍語抽出条件保持手段3は、入力される近傍語抽出条件を保持する。近傍語抽出手段4は、検索キーワード及び近傍語抽出条件に基づいて、文書データから近傍語を抽出する。近傍語リスト生成手段5は、近傍語リストを生成する。近傍語リスト保持手段6は、生成された近傍語リストを保持する。キーワード追加手段7は、近傍語リストから選択された近傍語を検索キーワードとして追加する。そして検索式作成手段8は、検索キーワードから検索式を作成する。

【0018】このような検索式作成装置にて文書検索のための検索式を作成する場合、ユーザが入力する検索キーワードは、検索キーワード保持手段2に保持される。すなわち、検索キーワード保持手段2は、入力される検索キーワードが「aaaa」ならば、これを保持する。また同様に、ユーザが入力する近傍語抽出条件は、近傍語抽出条件保持手段3に保持される。

【0019】そして近傍語抽出手段4は、ユーザの指示を受けると、検索キーワード及び近傍語抽出条件に基づいて、文書データ保持手段1に保持された文書データから、近傍語を抽出する。例えば検索キーワード「aaaa」を含む文書データとして「aaaa、bbbb、・・・」及び「aaaa、cccc、・・・」が保持されている場合、近傍語抽出条件に合致することを確認してから、「bbbb」や「cccc」を近傍語として抽出する。

【0020】この際、抽出元となる文書データのID等を抽出された近傍語に関連付けておくと、複数の文書データから同じ近傍語が抽出される場合もあり、近傍語と抽出元文書データとの関係は1:1とは限らない。なお、本明細書中における近傍語とは、検索キーワードの近傍（範囲は近傍語抽出条件にて定められる）に存在し、単独で意味をなす自立語を指す。

【0021】近傍語リスト生成手段5は、抽出された近傍語にて近傍語リストを生成する。すなわち、抽出された近傍語が「bbbb」と「cccc」であれば、これらをリストアップする。近傍語リスト保持手段6は、生成された近傍語リストを保持する。表示される近傍語リストを見たユーザが、リスト中の任意の近傍語を選択すると、キーワード追加手段7はこの選択を受け付け、選択された近傍語を、検索キーワード保持手段2に検索キーワードとして追加する。ここでユーザが近傍語「bbbb」を選択すれば、この単語が新たな検索キーワード

として追加される。

【0022】そして検索式作成手段8は、ユーザの指示を受けると、検索キーワード保持手段2に保持された検索キーワードから検索式を作成する。すなわち、この時点で検索キーワードとして保持されているのが「a a a a」と「b b b b」であれば、これらの単語から検索式を作成する。

【0023】このようにして検索式を作成すると、検索要求を絞り込むための情報、すなわち近傍語の提示が効率良く行われ、この近傍語を検索式を作成するための検索キーワードとして追加できるので、検索式作成にかかるユーザの負担を軽減することが可能となる。

【0024】ここで、本発明の検索式作成装置を文書検索装置と組み合わせて実施する場合の構成及び動作について説明する。図2は、本発明の検索式作成装置を適用した文書検索装置のハードウェア構成図である。

【0025】本発明の検索式作成装置を適用した文書検索装置10は文書情報記憶部20を持ち、入力装置30を介してユーザからの入力を受け付け、出力装置40を介してユーザに情報を出力する。ここで、入力装置30としてはマウス等のポインティングデバイスやキーボード等が適用可能である。また、出力装置40としてはモニタやプリンタ等が適用可能である。

【0026】文書検索装置10は、テキスト検索処理部11と、検索結果記憶部12と、近傍語抽出処理部13と、近傍語記憶部14と、同一文内近傍語検出処理部15と、品詞分類処理部16と、から構成されている。

【0027】ここで、テキスト検索処理部11はCPU等で構成され、入力装置30を介して入力される検索キーワードや検索式に基づいて、文書情報記憶部20から文書情報を検索する。検索にあたっては全文検索等の公知技術が適用される。

【0028】検索結果記憶部12はメモリ等で構成され、テキスト検索処理部11の検索結果を記憶する。すなわち、検索キーワードや検索式と合致した文章の文番号及びその文章を含む文書のID等を関連付けて記憶する。また、ここに記憶される検索結果は、出力装置40を介して出力される。

【0029】近傍語抽出処理部13はCPU等で構成され、入力装置30を介して入力される近傍語抽出条件に基づいて、検索結果記憶部12に記憶されている検索結果の文書データから近傍語を抽出する。近傍語を抽出するには、検索キーワードや検索式と合致した文章を形態素解析し、近傍語抽出条件に合った単語を検出すればよい。

【0030】なお、近傍語抽出条件は「[m] 検索キーワード[n]」といった形式で入力される。これは、検索キーワードの前方に位置し、単独で意味を持つ自立語m個と、検索キーワードの後方に位置し、単独で意味を持つ自立語n個とを、近傍語として抽出することを示

す。また、「検索キーワード1\*[m] 検索キーワード2[n]」（\*は論理積）と入力された場合には、検索キーワード1と検索キーワード2との両方を含む文書を対象として、「[m] 検索キーワード2[n]」を満たす近傍語を抽出すればよい。

【0031】また、この際行われる形態素解析は自然言語処理の基本技術として広く知られる技術で、「自然言語処理の基礎技術」（野村浩郷著、社団法人電子情報通信学会、1988）等に詳しい。

【0032】近傍語記憶部14はメモリ等で構成され、近傍語抽出処理部13にて抽出される近傍語を記憶する。なお、近傍語は、その近傍語を抽出した元の文章の文番号及び文書のID等と関連付けて記憶する。また、ここに記憶された近傍語は品詞分類処理部16に入力される。

【0033】同一文内近傍語検出処理部15は、出力装置40を介して出力された近傍語のリストからユーザが任意の近傍語を選択して検索キーワードを追加した場合に、追加された検索キーワードを含む文を近傍語記憶部14で検索し、この文に含まれる近傍語を検出する。検出した近傍語は品詞分類処理部16に入力される。

【0034】品詞分類処理部16は、入力された近傍語を品詞、すなわち、名詞、動詞・・・等に分類して近傍語リストを生成する。生成した近傍語リストは出力装置40を介して出力する。

【0035】次に、この文書検索装置10にて文書検索を行う手順を説明する。図3は、本発明の検索式作成装置を適用した文書検索装置にて文書を検索する手順を示したフローチャートである。以下、ステップ番号に沿って説明する。

【S1】テキスト検索処理部11は、入力装置30を介して入力される検索キーワードを受け付ける。

【S2】テキスト検索処理部11は、受け付けた検索キーワードにて文書情報記憶部20を検索し、検索キーワードを含んだ文章の文番号及びその文章を含む文書データのID等を検索結果として検索結果記憶部12に記憶させる。

【S3】近傍語抽出処理部13は、入力装置30を介して入力される近傍語抽出条件を受け付ける。

【S4】近傍語抽出処理部13は、入力された検索キーワードもしくは検索式を含んだ文章を1つ選択し、この文章を形態素解析する。なお、この形態素解析の際、名詞が連続して出現した場合には、その部分を単独の複合語として扱うこととする。よって、例えば「形態素」という単語と「解析」という単語とが連続して出現した場合には、「形態素解析」という単語が出現したものとして扱う。

【S5】近傍語抽出処理部13は、検索キーワード及び近傍語抽出条件に基づいて、形態素解析した文章から近傍語を抽出し、その文章の文番号及びその文章を含む文

10

20

30

40

50

書のID等を、近傍語と関連付けて近傍語記憶部14に記憶させる。

【S6】近傍語抽出処理部13は、検索結果記憶部12にステップS4～S5の処理の済んでいない文書データがあるか否かを判断する。未処理の文書データがあればステップS4へ再度進み、未処理の文章を選択する。また、未処理の文書データがない場合はステップS7へ進む。

【S7】近傍語記憶部14は、記憶した近傍語をすべて品詞分類処理部16に入力する。品詞分類処理部16は、この近傍語を品詞分類して近傍語リストを生成する。

【S8】品詞分類処理部16は、生成した近傍語リストを出力装置40を介して出力する。

【S9】同一文内近傍語検出処理部15は、入力装置30を介してユーザから、追加する検索キーワードが選択されたか否かを判断する。ここで選択されるのは、ステップS8で表示した近傍語リストに含まれた近傍語の1つである。追加する検索キーワードが選択された場合はステップS10に進む。また、追加する検索キーワードが

選択されない場合はステップS11に進む。

【S10】検索キーワードが複数存在するので、検索式を作成する。この詳細については後に説明する。検索式が作成した後、ステップS4に再度進む。

【S11】ここまでの処理で作成された検索式があれば、その検索式はテキスト検索処理部11に入力される。テキスト検索処理部11は、その検索式に基づいて文書情報記憶部20を検索する。検索結果は検索結果記憶部12に記憶される。なお、検索式が作成されていない場合には、この処理を省いてもよい。

【S12】検索結果記憶部12は、記憶した検索結果を出力装置40を介して出力する。

【0036】ここで、ステップS10で行う検索式の作成について説明する。出力装置40としてモニタ等の表示装置が適用されている場合、この時点でこの表示装置には近傍語のリストが表示されている。

【0037】図4は、近傍語リストの表示画面例を示す図である。表示画面50には、近傍語リスト51と、開始ボタン56と、終了ボタン57とが表示されている。

【0038】ユーザは、検索キーワードを追加して同一文内の近傍語を検出したい場合に、マウス等のポインティングデバイスで開始ボタン56を選択する。また、検索キーワードの追加の必要がない場合に、終了ボタン57を選択する。

【0039】近傍語リスト51は、近傍語抽出条件表示欄52と、近傍語<名詞>表示欄53と、近傍語<動詞>表示欄54と、近傍語<その他>表示欄55と、から構成されている。また、近傍語<名詞>表示欄53及び近傍語<動詞>表示欄54には、それぞれスクロールバーが付随している。

【0040】近傍語抽出条件表示欄52には、入力された近傍語抽出条件が表示される。また、近傍語<名詞>表示欄53には、抽出された近傍語のうち、名詞が表示される。近傍語<動詞>表示欄54には、抽出された近傍語のうち、動詞が表示される。そして、近傍語<その他>表示欄54には、抽出された近傍語のうち、名詞でも動詞でもないものが表示される。

【0041】なお、ここでは検索キーワードを「規制」、近傍語抽出条件を「[2]規制」として文書検索を行った場合の例を示している。これは、近傍語抽出条件「[2]規制[0]」と等価である。

【0042】検索キーワード「規制」を図2に示した入力装置30を介して入力した場合、テキスト検索処理部11は、文書情報記憶部20に記憶されている文書データに対し、この検索キーワード「規制」に基づいた文書検索を行う。この検索結果は検索結果記憶部12に記憶される。

【0043】近傍語抽出条件「[2]規制」を入力装置30を介して入力された近傍語抽出処理部13は、検索結果記憶部12に記憶された検索結果の文書データを形態素解析する。そして、検索キーワード「規制」の出現の度に、この検索キーワード「規制」より前方に出現し、単独で意味を持つ単語を2つ抽出し、近傍語とする。抽出した近傍語は近傍語記憶部14に記憶され、品詞分類処理部16にて品詞分類されて近傍語リスト51が生成される。

【0044】このように、抽出される近傍語を品詞別に分類することで、ユーザは検索キーワードを追加するにあたって、適当な単語を探しやすくなる。なお、ここでは検索キーワードの前方に出現する単語のみを近傍語とする場合を説明したが、近傍語抽出条件によって検索キーワードの後方に出現する単語を近傍語として抽出する場合もある。その場合には、前方に出現した近傍語と後方に出現した近傍語とを区別して表示させてもよい。

【0045】また、ここでは近傍語を品詞別に分けて表示したが、これを分けずに表示することも可能である。次に、ユーザが検索キーワードを追加する場合を説明する。

【0046】「譲渡、取引、輸出入に関する規制」に関心を持った場合、ユーザはマウス等のポインティングデバイスで近傍語「譲渡」、「取引」、「輸出入」を指定してから、開始ボタン56を選択する。

【0047】図5は、追加する検索キーワードを指定し、同一文内の近傍語を検出した様子を示す図である。表示画面50aには、近傍語リスト51aが表示される。

【0048】近傍語記憶部14では先に近傍語抽出条件「[2]規制」に基づいて近傍語を抽出した際に、各近傍語と、その近傍語を抽出した元文書のID等とを対応付けて記憶しているので、検索キーワードを追加される

と同一文内近傍語検出処理部15は、近傍語記憶部14から同一文内に存在する近傍語を検出する。

【0049】従って検索キーワードとして「譲渡」、「取引」、「輸出入」が追加指定されると、「規制」という単語を含み、なおかつ「譲渡」もしくは「取引」もしくは「輸出入」という単語を含む文書から、「規制」という単語の前方に出現し、単独で意味を持つ単語2つが近傍語として検出され、近傍語リスト51aに表示される。近傍語リスト51aに提示されている近傍語の数は、近傍語リスト51に提示されている近傍語の数と比較すると少なくなっている。

【0050】ここでユーザがさらに「野生生物、希少野生動物の譲渡、取引、輸出入に関する規則」に関心を持った場合、ユーザは近傍語「野生生物」、「希少野生動物」を指定してから、開始ボタン56を選択する。

【0051】図6は、追加する検索キーワードを指定し、同一文内の近傍語を検出した様子を示す図である。表示画面50bには、近傍語リスト51bが表示される。

【0052】検索キーワードとして「野生生物」、「希少野生動物」が追加指定されると、「規制」という単語を含み、「譲渡」もしくは「取引」もしくは「輸出入」という単語を含み、なおかつ「野生生物」もしくは「希少野生動物」という単語を含む文書から、「規制」という単語の前方に出現し、単独で意味を持つ単語2つが近傍語として検出され、近傍語リスト51bに表示される。近傍語リスト51bに提示されている近傍語の数は、近傍語リスト51aに提示されている近傍語の数と比較するとさらに少なくなっている。

【0053】従って、この時点で、文書情報記憶部20には検索式「規制\*(譲渡+取引+輸出入)\*(野生生物+希少野生動物)」(\*は論理積を、+は論理和を示す)を満たす文書が記憶されていることが判る。ユーザは、このように近傍語の数が十分に減ったと判断した時点で、表示されている近傍語抽出条件から検索式を作成すればよい。

【0054】この検索式は、近傍語抽出条件表示欄52、52a、52b・・・に表示されている文字列の[m]及び[n]を削除し、「:」を「\*」に、「[」を「+」に変換すれば作成できるので、CPU等に処理させてもよいし、ユーザが入力してもよい。

【0055】このように、本発明の検索式作成装置を適用した文書検索装置10では、検索キーワードに対し近傍語が抽出、表示され、ユーザが関心のある近傍語を次々に絞り込んで指定することができるので、複雑な操作

を必要とせずに適切な検索式を作成することが可能である。

【0056】また、上記で説明したように、近傍語を段階的に選択していくことができるので、入力する検索キーワードは、広い概念を持つものから順次、狭い概念を持つものに絞り込んでいくことができるので、ユーザ自身が検索意図を掴めていないような場合にも、必要な文書の検索が可能となる。

【0057】なお、図4～図6で示した例では、図5から図6へ移った時点で、「輸出入」という近傍語が消えている。このことから、「輸出入」という単語が、「野生生物」や「希少野生動物」という単語と、近傍語として両立している文書が存在しないことが判る。よって、この場合の検索式は「規制\*(譲渡+取引)\*(野生生物+希少野生動物)」でもよい。

【0058】

【発明の効果】以上説明したように本発明の検索式作成装置は、検索要求を絞り込むための情報、すなわち検索キーワード及び近傍語抽出条件に基づいて抽出される近傍語の提示を効率良く行い、この近傍語を検索式を作成するための検索キーワードとして追加できる構成としたので、検索式作成にかかるユーザの負担を軽減することが可能となる。

【図面の簡単な説明】

【図1】本発明の検索式作成装置の原理構成を示したブロック図である。

【図2】本発明の検索式作成装置を適用した文書検索装置のハードウェア構成図である。

【図3】本発明の検索式作成装置を適用した文書検索装置にて文書を検索する手順を示したフローチャートである。

【図4】近傍語リストの表示画面例を示す図である。

【図5】追加する検索キーワードを指定し、同一文内の近傍語を検出した様子を示す図(その1)である。

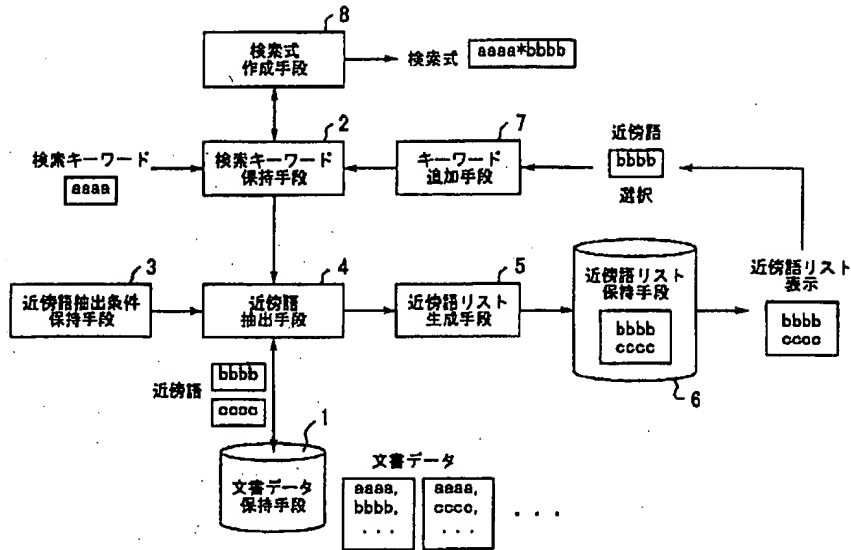
【図6】追加する検索キーワードを指定し、同一文内の近傍語を検出した様子を示す図(その2)である。

【符号の説明】

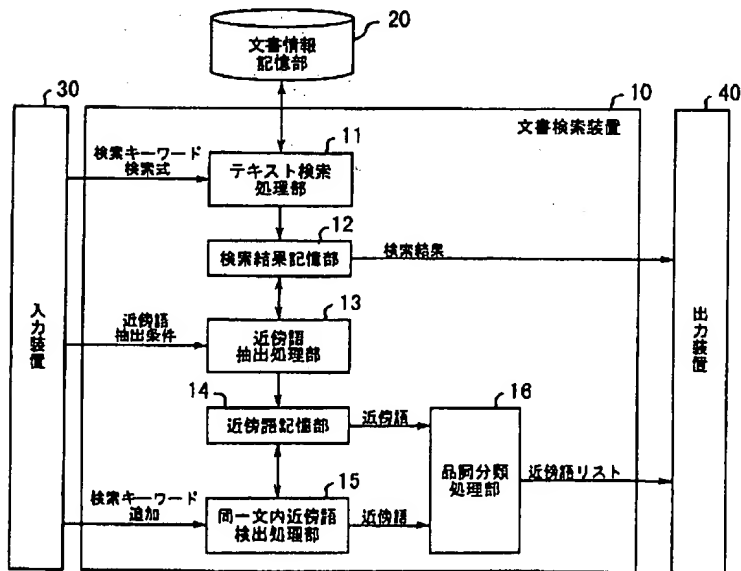
- 1 文書データ保持手段
- 2 検索キーワード保持手段
- 3 近傍語抽出条件保持手段
- 4 近傍語抽出手段
- 5 近傍語リスト生成手段
- 6 近傍語リスト保持手段
- 7 キーワード追加手段
- 8 検索式作成手段



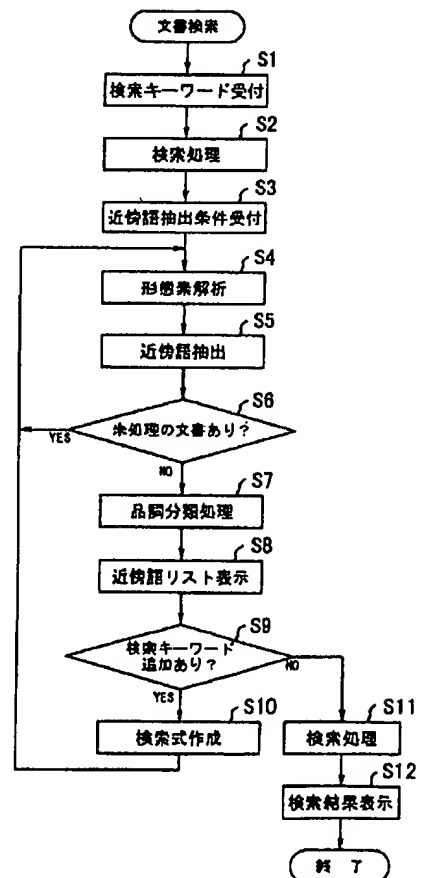
【図1】



【図2】



【図3】



【図4】

50 表示画面

51 近傍語リスト

56 開始

52

53

【2】規制

<名詞>

硫黄酸化物  
大阪市  
環境  
観点  
希少野生動物  
特定地域  
健康保護  
行為  
航空機  
国際  
水質汚濁性農業  
大都市圏総量  
炭化水素  
特定有害廃棄物  
排水  
揚ガス  
野生生物

54

<動詞>

向上  
譲渡  
取引  
譲出  
売買  
販売  
輸出入

55

<その他>

厳しい

終了

57

【図5】

50a 表示画面

51a 近傍語リスト

56 開始

52a

【2】規制：(譲渡, 取引, 輸出入)

53a

<名詞>

硫黄酸化物  
希少野生動物  
国際  
特定有害廃棄物  
野生生物

54a

<動詞>

譲渡  
取引  
売買  
販売  
輸出入

55a

<その他>

終了

57

【図6】

50b 表示画面

51b 近傍語リスト

56 開始

52b

【2】規制：(譲渡, 取引, 輸出入)：(野生生物, 希少野生動物)

53b

<名詞>

希少野生動物  
野生生物

54b

<動詞>

譲渡  
取引

55b

<その他>

終了

57